WHAT IS CLAIMED IS:

1. An apparatus for expanding a character string, wherein the character string is entered to search image information of documents, the apparatus comprising:

a character string dividing device to divide the entered character string into a plurality of partial character strings each having a plurality of characters;

a referencing device to reference a similarity table, the similarity table previously storing groups of similar partial character strings, each of the groups of similar partial character strings being derived from each of the plurality of partial character strings obtained from the character string dividing device by changing at least one of the characters of each partial character string to a different character which is similar in shape; and

an expansion device to combine the plurality of similar partial character strings given by the referencing device into expanded words and store them in an expanded word table.

2. The apparatus according to claim 1, wherein the similarity table is arranged in the order of their emergence probability in each group and has only those similar partial character strings whose emergence probabilities are greater than a predetermined value.

3. The apparatus according to claim 1, wherein,

when the similarity table does not include similar different characters, the referencing device gives the partial character strings obtained from the character string diving device to the expansion device, and the expansion device uses the partial character strings to produce the expanded words.

4. The apparatus according to claim 1, wherein, when the similarity table does not have entries for the partial character strings obtained from the character string diving device, the referencing device references a second similarity table which stores in advance second groups of similar partial character strings arranged in the order of magnitude of their emergence probability in each group, each of the second groups of similar partial character strings being derived from each of short partial character strings made up of a smaller number of characters than the partial character strings obtained from the character string diving device by changing at least one of the characters of each short partial character string to a different character which is similar in shape.

5. The apparatus according to claim 1, wherein, when the entered character string is not divisible into the plurality of partial character strings without a remainder, characters adjoining each character of the remainder character string are added to the each character so that resultant character strings have the same number of characters as the divided character

strings, and the character strings thus obtained are added to the plurality of partial character strings.

6. In a system for retrieving a document containing a search character string specified by an operator in a search text documents that are produced by performing character recognition processing on image documents, a search character string expanding method comprising:

a search character string dividing step of dividing the entered search character string into partial character strings each consisting of a predetermined number n of characters ($n \geq 2$);

a similarity table referencing step of checking the n-character partial character strings ($n \geq 2$) against an n-character-based similarity table, the n-character-based similarity table being generated in advance by storing character strings of similar character shapes that are highly likely to be erroneously recognized; and

a search character string expanding step of extracting groups of similar character strings by checking the partial character strings making up the search character string against the n-character-based similarity table and combining the extracted similar character strings to generate expanded words.

7. A search character string expanding method according to claim 6, wherein entry characters in the n-character-based similarity table include only a part

of partial character strings each of which is a combination of n characters.

8. A search character string expanding method according to claim 7, wherein when a partial character string making up the search character string is not found in the n-character-based similarity table, similar character strings to the partial character string are not extracted.

9. A search character string expanding method according to claim 7, wherein when a partial character string making up the search character string is not found in the n-character-based similarity table, an m-character-based similarity table, which is prepared in advance by storing similar m-character strings ($m < n$) of similar character shapes highly likely to be erroneously recognized, is referenced to generate expanded words.

10. A search character string expanding method according to claim 6, further including a expansion method switching step of calculating a length of the search character string and selecting between expanded word generation methods according to the search character string length.

11. In a system for retrieving a document containing a search character string specified by an operator in a search through text documents that are produced by performing character recognition processing on image documents, a search character string expanding

method comprising:

a expansion method switching step of calculating a length of the search character string and selecting between expanded word generation methods according to the search character string length.

12. A search character string expanding method according to claim 10, wherein the number of expanded character strings generated is adjusted according to the search character string length.

13. A search character string expanding method according to claim 11, wherein whether the expanded words are generated or not is determined according to the search character string length.

14. A search character string expanding method according to claim 13, wherein setting information is provided for selecting between the expanded word generation methods.

15. A document information retrieval method comprising:

a text search step of executing a search by using as a search condition a logical sum of expanded search character strings obtained by the search character string expansion method of claim 14.

16. A program read into and running on a computer to expand a character string, wherein the character string is entered to search image information of documents, the program comprising:

a character string dividing step of dividing

the entered character string into a plurality of partial character strings each having a plurality of characters;

a referencing step of referencing a similarity table, the similarity table previously storing groups of similar partial character strings, each of the groups of similar partial character strings being derived from each of the plurality of partial character strings obtained from the character string dividing step by changing at least one of the characters of each partial character string to a different character which is similar in shape; and

an expansion step of combining the plurality of similar partial character strings given by the referencing step into expanded words and store them in an expanded word table.

17. The program according to claim 16, wherein the similarity table is arranged in the order of their emergence probability in each group and has only those similar partial character strings whose emergence probabilities are greater than a predetermined value.

18. The program according to claim 16, wherein, when the similarity table does not include similar but different characters, the referencing step gives the partial character strings obtained from the character string diving step to the expansion step, and the expansion step uses the partial character strings to produce the expanded words.

19. The program according to claim 16, wherein, when the similarity table does not have entries for the partial character strings obtained from the character string diving step, the referencing step references a second similarity table which stores in advance second groups of similar partial character strings arranged in the order of magnitude of their emergence probability in each group, each of the second groups of similar partial character strings being derived from each of short partial character strings made up of a smaller number of characters than the partial character strings obtained from the character string diving step by changing at least one of the characters of each short partial character string to a different character which is similar in shape.

20. The program according to claim 16, wherein, when the entered character string is not divisible into the plurality of partial character strings without a remainder, characters adjoining each character of the remainder character string are added to the each character so that resultant character strings have the same number of characters as the divided character strings, and the character strings thus obtained are added to the plurality of partial character strings.